神经模拟推断：基于神经网络和模拟推断的认知建模方法

潘晚坷

胡传鹏

(南京师范大学心理学院，南京 21002)

摘要

认知建模是量化和理解人类心智过程的重要方法，但基于认知过程的计算模型往往较为复杂，难以确定其似然函数，进而导致模型参数估计及模型比较的困难。基于模拟的推断方法可以规避计算似然函数，直接利用模拟数据来估计模型参数或进行模型比较。尽管如此，该方法仍面临估计不准确和计算效率低的问题。神经网络技术，通过在训练阶段分摊生成模拟数据的成本，推动了基于神经网络的模拟推断方法的发展，从而提高了模型推断的准确性和效率。该方法已成功应用于近似认知模型的似然函数、估计参数的后验分布和近似的模型比较。然而，该类方法在训练成本效益、有效性验证以及方法和应用整合方面仍需进一步研究。

关键词 认知建模，生成模型，贝叶斯，基于模拟的推断，基于神经网络的模拟推断

基于神经模拟推断的认知建模方法

1. 引言

认知科学的核心在于构建能够解释大脑如何产生行为和认知功能的计算模型(Churchland & Sejnowski, 1988; McClelland, 2009)。计算模型有利于整合多模态数据，量化个体差异，并能为精神疾病提供诊断依据(Eaton et al., 2023; Geng et al., 2022; Kriegeskorte & Douglas, 2018; Wang & Krystal, 2014)。认知建模将认知理论实例化为计算模型，允许研究者通过调整模型参数来量化理论对于认知活动的预测。

认知建模中最为关键的一环是模型推断，通过估计参数值来量化认知过程，并为模型预测、验证和比较提供基础(Guest & Martin, 2021; Schad et al., 2021; Wilson & Collins, 2019)。尽管基于认知过程的计算模型具有多方面的优势，但在模型推断方面仍面临多重挑战（见表 1）。首先，认知计算模型通常容易被构建为生成模型，研究者可通过调整模型参数来生成模拟数据。然而，当尝试用模型似合实际数据并对模型参数进行估计时，通常涉及模型似然函数的解析数学推导(Kangasräasiö et al., 2019; Turner & Zandt, 2018)，这往往会导致模型难以求解。其次，在认知研究中，数据的层级结构通常要求使用层级模型同时估计群体水平与个体水平的参数，这使得模型参数的数量随被试和实验条件成线性增长，进一步加剧了模型参数估计的复杂性(Boehm et al., 2023; Vandekerckhove et al., 2011)。最后，即便能够完成参数估计，评估模型的有效性以及在多个模型中选择最优模型，目前仍极具挑战性(Gelman et al., 2014; Gronau et al., 2017)。

表 1 认知建模优势和挑战

	生成式认知模型	模型分层化	模型比较和平均
问题来源	认知建模将认知理论或认知机制实例化为生成式计算模型	心理学相关领域通常涉及层级化数据结构，通常包含同一被试在多个条件下的重复测试	不同模型具有特定的假设、适用条件及其实现方法
挑战	对于生成模型的推断涉及逆问题，需要将生成模型转化为具有解析形式的目标函数(例如似然函数)，而对于复杂的认知模型，似然函数的推导颇具挑战性	随着被试数量的增加，分层模型参数总量也相应线性增长，使得对分层模型的拟合和推断变得极具挑战性	即使具有似然函数，边际似然通常也难以计算，使得模型评估困难；不同参数估计方法下的模型评估存在差异

针对以上这些问题，研究者采用贝叶斯方法对参数后验进行估计。得益于概率编程语言的出现和马尔可夫链蒙特卡洛采样方法(MCMC)的发展，研究者无需自己解决后验采样

算法的问题，因此，贝叶斯方法在认知模型中获得广泛应用(Kruschke, 2014; Ma, 2022; Schad et al., 2021)。

然而，MCMC 方法无法解决缺乏显性似然函数的认知模型。为解决这一问题，研究者绕过似然函数的计算，通过生成模拟数据来获得参数样本，这类方法即基于模拟的推断(Simulation-based Inference, SBI)，简称为模拟推断。模拟推断已广泛应用于多个学科领域(Cranmer et al., 2020)，并在认知建模领域也获得了一定的应用(Turner & Zandt, 2018)。尽管如此，模拟推断方法仍然受到数据和参数维度的限制，在处理大规模数据集和高度复杂的模型时面临着计算效率低下的问题。最近，神经网络和深度学习方法被引入到模拟推断中，旨在克服处理高维数据、计算资源限制以及适应多样化的应用场景(Cranmer et al., 2020)。

神经网络和深度学习方法，不仅能够模拟大脑的复杂神经网络结构，还能够通过学习数据中的复杂模式来增强模型的预测能力(Gonçalves et al., 2020; Hashemi et al., 2023)。例如，神经网络作为函数逼近器，在处理高维数据和参数时具有优势，可以用于近似无法通过解析方式计算的似然函数(Boelts et al., 2022; Fengler et al., 2021)、估计认知模型的参数或进行模型比较(Radev, Voss, et al., 2020; Radev, Schmitt, Pratz, et al., 2023)。正因如此，“基于神经网络的模拟推断”(Neural Simulation-based Inference, NSBI，简称“神经模拟推断”)在认知建模领域中开始产生如下三类应用：神经似然估计、神经后验估计以及神经模型比较。这些应用将有望进一步扩大计算模型在认知科学中的应用范围。

2. 模拟推断方法

对于认知模型而言，即便模型具备明确的似然函数，模型参数的后验分布也可能不具备封闭形式的解。因此，研究者采用 MCMC 和变分推断(variantional inference, VI)等进行模型推断(Farrell & Lewandowsky, 2018; Forstmann & Turner, 2024)。但是当模型的似然函数不明确或难以计算时，研究者可能不得不放弃复杂的模型，或者采用避免计算似然函数的方法。近年来，研究者通过模拟器生成模拟数据来对真实观测数据进行推断，这种方法被称为无似然推断(likelihood-free inference)技术，也被称为基于模拟的推断(Cranmer et al., 2020)。当前的模拟推断有两种策略（见图 1A）：近似贝叶斯计算和概率密度近似。

近似贝叶斯计算(Approximate Bayesian Computation，ABC)是针对贝叶斯模型设计的框架，已成为解决无似然推理问题的重要工具(Kangasrääsiö et al., 2019; Palestro et al., 2018; Turner & Zandt, 2018)。近似贝叶斯的核心思想在于通过生成和处理模拟数据，间接逼近难以直接计算的后验分布$p(\theta|D)$。其步骤包括：从先验分布中抽取参数样本；用这些参数生成模拟数据；计算模拟数据和实际数据的摘要统计量；比较这些统计量的差异；根据预设

的阈值决定是否接受参数作为后验样本。重复这一过程，累积足够的后验样本，以此构建后验分布的近似(Palestro et al., 2018)。

部分研究者采用概率密度近似(Probablistic Density Approximation, PDA)以替代近似贝叶斯计算中的摘要统计，以实现更准确的参数估计(Holmes, 2015; Turner & Sederberg, 2014)。概率密度近似的核心思想在于利用整个模拟数据的分布信息来近似似然函数，称为伪似然(pseudo-likelihood)，记作$\hat{L}(\theta^*|\tilde{D})$。该方法需要先根据参数生成大量模拟数据，以覆盖该参数下似然分布，随后使用核密度估计方法来加权平均估计模拟数据的概率密度函数(Turner & Sederberg, 2014)。概率密度近似具有似然函数的优势，既适用于频率优化方法，也能与 MCMC 方法结合，用于可计算模型评估指标 (Holmes, 2015; Robert et al., 2011)。

然而，近似贝叶斯计算和概率密度近似都面临着计算成本高昂的挑战，尤其是在处理高维数据或复杂模型时，所需的模拟数据量会呈指数级增长（即“维度灾难”）。这意味着，当模拟数据量不足以充分探索参数空间时，近似计算的准确度会受到影响。

3. 基于神经网络的模拟推断方法

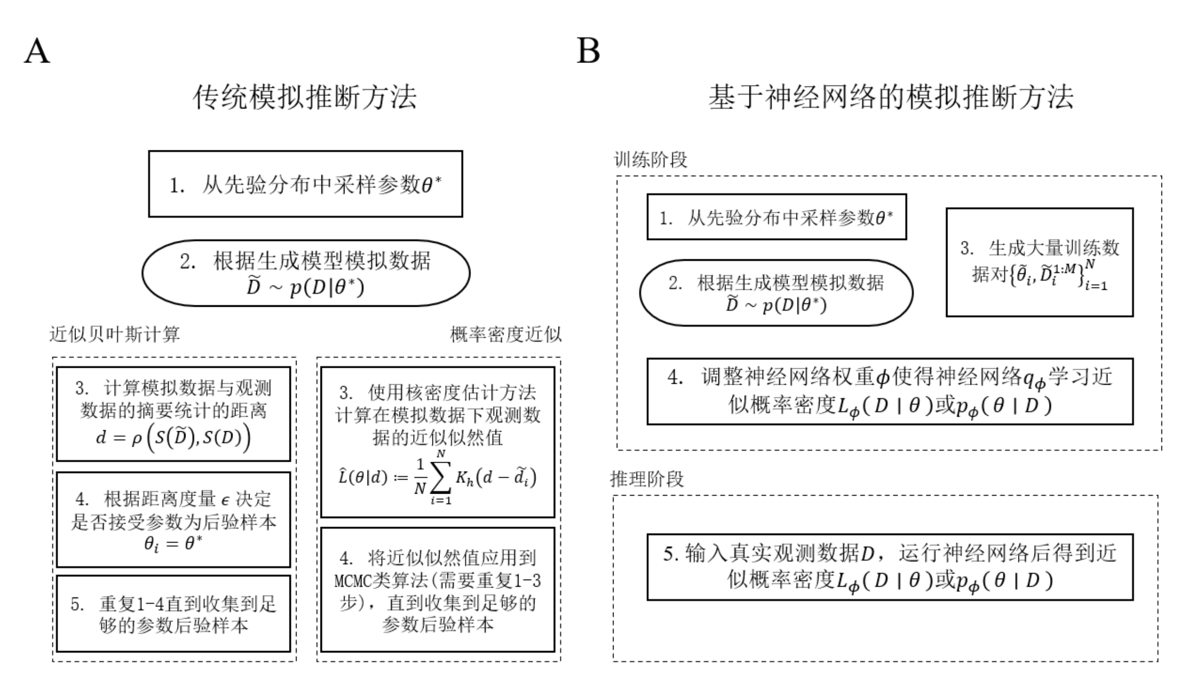


图 1 **传统模拟推断方法和基于神经网络的模拟推断方法的对比**。椭圆形图示代表的是模拟数据生成的过程，这一环节通常消耗最多的计算资源。与传统的模拟方法相比，采用基于神经网络的模拟推理方法能在训练阶段分散大量模拟数据的计算需求，因此在推理阶段无需再生成模拟数据即可进行推理，显著提高了效率。

为了克服传统模拟推断方法在计算资源、处理高维数据以及适用场景方面的挑战，研究者尝试引入机器学习和深度学习技术。例如，将随机森林(random forest, RF)或深度神经网络(deep neural network, DNN)和近似贝叶斯计算结合，研究者能够有效处理高维数据问题，提高推理的质量(Radev, Mertens, et al., 2020; Raynal et al., 2019)。

神经网络和深度学习模型的优势在于具有分摊(amortisation)特性(见图 1B)，使得生成模拟数据的成本被分摊到训练阶段，一旦训练完成，神经网络模型便能实现快速推理，而无需额外生成模拟数据(Radev et al., 2022; Radev, Schmitt, Pratz, et al., 2023)。此外，最近，神经密度估计领域的发展推动了深度神经网络在直接估计条件概率密度方面的应用，提供了一种替代传统模拟推断方法的新途径 (Liu et al., 2021; Papamakarios et al., 2017; Papamakarios & Murray, 2016)。例如，神经密度估计技术能够用于准确估算贝叶斯公式中的后验概率分布和似然函数，从而增强统计推断的效率和准确性(Brehmer et al., 2020; Mancini et al., 2023)。

本文将神经网络和深度学习应用于模拟推断统的方法称为神经模拟推断，并根据其应用领域分为三个类别(见表 2)，包括神经似然估计、神经后验估计，以及基于神经网络的模型比较。该分类方式有助于理解不同方法在认知模型推断中的具体作用和应用场景。本文将详述三种方法在认知建模领域的发展、原理，并全面评析其优缺点和应用范围。

表 2 神经模拟推断的应用比较

	神经似然估计	神经后验估计	神经模型比较
描述	根据输入的先验参数和给定的观测数据输出似然值 $L_\phi(D \mid \theta)$	根据给定观测数据输出参数后验分本的样本 $p_\phi(\theta \mid D)$	根据输入的数据给出对应的模型证据 $p_\phi(\mathcal{M}_j \mid D)$
优点	可结合 MCMC 等算法进行参数估计； 并且可以结合后验参数和桥采样等算法用于计算模型比较评估指标; 可用于分层模型	无需进行 MCMC 采样，能快速获得参数后验分布样本	通过单个神经网络模型考虑多个认知模型的比较； 无需重新训练，可通过调试神经网络扩展到新的认知模型
缺点	受到优化算法(例如 MCMC)的限制； 模型推理速度较慢	无法获得似然值，无法直接进行模型评估和比较； 需根实验设计训练专门的神经网络	无法对模型参数进行推断
应用示例	(LANs, Fengler et al., 2021) (MNLE, Boelts et al., 2022) (JANA, Radev, Schmitt, Pratz, et al., 2023)	(Bayesflow, Radev, Voss, et al., 2020) (JANA, Radev, Schmitt, Pratz, et al., 2023) (ANN, Rmus et al., 2024)	(DEN Radev et al., 2021) (DHMC, Elsemüller et al., 2023) (ANN, Rmus et al., 2024)
工具	(LANs, Fengler et al., 2021) (sbi, Tejero-Cantero et al., 2020) (Bayesflow, Radev, Schmitt, Schumacher, et al., 2023)	(sbi, Tejero-Cantero et al., 2020) (Bayesflow, Radev, Schmitt, Schumacher, et al., 2023)	(sbi, Tejero-Cantero et al., 2020) (Bayesflow, Radev, Schmitt, Schumacher, et al., 2023)

3.1. 神经似然估计. 在传统的模拟推理方法中，虽然通过密度概率近似可以获得似然值，但该方法受到维度诅咒的限制，难以处理高维数据，并且生成大量数据的成本高昂。相比之下，神经网络方法在处理高维数据方面具有显著优势，且其分摊的性质使得生成大量模拟数据的过程被分摊在训练阶段，因此在训练完成后可以快速获得似然值。

鉴于此，神经似然估计方法被引入认知建模中，作为概率密度近似的替代模型，其目标在于免去每次计算似然值时需重复生成模拟数据的过程，同时所计算的似然值可用于频率推断或者 MCMC 等方法。Fengler 等人(2021)提出"似然近似网络"(Likelihood Approximate Networks, LANs)，以快速和准确地估计认知模型的似然函数。

LANs 采用多层感知机(Multilayer Perceptrons, MLPs)来学习参数和数据到对数似然函数的映射 $f_{\mathrm{MLP}(\phi)}: (\theta, D) \to \log L\,(D|\theta)$，其中$f_{\mathrm{MLP}(\phi)}$代表具有权重参数$\phi$的多层感知机神经网络。LANs 执行监督学习任务，其中输入为参数$\tilde{\theta}$和数据$\tilde{D}$，输出为对数似然函数值$\log \tilde{L}\,(\tilde{D}|\tilde{\theta})$。通过优化网络权重$\phi$以最小化对数似然估计$\log \tilde{L}\,(\tilde{D}|\tilde{\theta})$与实际对数似然值（标签数据）$\log L\,(\tilde{D}|\tilde{\theta})_i^{1:M}$之间的差异，最终习得数据到对数似然函数的映射。

训练后的 LANs 可以替代似然函数，与现有的 MCMC 法结合使用 (Fengler et al., 2021)。LANs 的显著优势在于能够评估单个观测值的点似然性(pointwise likelihood)，这使其适用于分层模型和分层线性模型。尽管 LANs 能显著提升对于复杂认知模型推断的能力，但仍然存在局限性。首先，LANs 的训练成本高昂，需要数十亿次模拟(Fengler et al., 2021)，这限制了其广泛应用，因为大多数用户无法实施如此大规模的训练。此外，LANs 并非直接进行密度估计，而是采用类似于回归模型的方法建立数据和参数到似然值的映射。这种模型架构一方面要求大量参数以覆盖参数空间，另一方面，即便能够覆盖，也难以在参数空间中实现平滑的估计(Boelts et al., 2022)。

为了克服 LANs 的局限性，Boelts 等(2022)提出了混合神经似然估计(Mixed Neural Likelihood Estimation, MNLE)。MNLE 采用标准化流(Normalizing Flow, Papamakarios, 2019; Papamakarios et al., 2021)，可实现对条件密度的直接估计。不同于 LANs 使用类似"回归"的方式建立数据和参数到似然值的映射，MNLE 直接估计似然函数，且其所需的训练数据更少(Boelts et al., 2022)。与 LANs 相比，MNLE 具有两点优势：首先，MNLE 采用生成模型构建执行无监督学习任务，因此 MNLE 不再依赖似然值作为标签；其次，MNLE 中的参数和数据点之间是一一对应的关系，而无需根据每组参数生成大量数据点，因此其效应更高。Boelts 等(2022)的研究结果显示，MNLE 仅需万级别的数据量即可达到或超越 LANs 基于百亿模拟数据所达到的准确性。MNLE 已被整合至 Python 开源工具包"sbi"中，这为在认知神经科学领域中实现无似然推断方法提供了基础(Tejero-Cantero et al., 2020)。

LANs 和 MNLE 表明神经似然估计在认知建模领域的可行性。然而，这类方法仍有改进的空间。首先，神经似然估计在实际运用时仍然需要使用优化或近似方法，如 MCMC，这导致其推断效率仍然不高。其次，如果神经网络没有经过充分的训练，网络本身的噪声加上近似方法的噪声会进一步放大估计的不准确性。此外，尽管神经似然估计通常能利用图形处理单元(GPU)加速推断的过程，但当需要并行运行多条 MCMC 链时，这可能会与 GPU 的架构不匹配，从而降低其计算效率。但这一类方法的优势在于：如果某个模型较为常见，则可以由开发者进行训练后纳入到已有的工具包（如 HSSM）之中，用户直接调用该神经网络作为似然即可。

3.2. 神经后验估计. 神经似然估计对参数的推断是间接的，需要与 MCMC 等方法相结合，而神经后验估计则专注于直接对参数的后验分布进行估计。神经后验估计可以被视为基于深度学习的近似贝叶斯计算的改进，旨在解决后者面临的两个核心挑战：一是推断过程中倾向于估计后验分布的摘要统计量而非完整的后验密度；二是在处理数据时，仅能针对预设数量的观测值进行参数估计，缺乏对数据量的适应性。

Radev 等(2020)首次将神经后验估计方法应用于认知建模领域，提出了 BayesFlow 的框架(Radev et al., 2022; Radev, Voss, et al., 2020)。类似于 MNLE，BayesFlow 也采用基于标准化流的神经网络架构。在这种架构中，认知模型的参数后验分布 $p(\theta \mid D)$可以表示为在给定数据D的条件下参数θ和潜在变量z_θ(来自于简单分布)之间的标准化流$p_\phi\left(\theta \mid \mathcal{H}_\eta(D)\right) = p(z_\theta)\left|\det\left(\frac{\partial f^{-1}}{\partial \theta}\right)\right|$。与 MNLE 的不同之处在于，BayesFlow 引入摘要统计网络(summary statistics network)，即$\mathcal{H}_\eta$，以解决基于深度学习的近似贝叶斯计算不允许试次数量变化的问题，该网络可以将任意长度的数据$D^{1:M}$(假设这些数据来自于独立同分布，并且具有置换不变性)映射为固定长度的嵌入向量$\mathcal{H}_\eta(D)^{1:K}$，其中$K$代表摘要统计维度的潜在维度。

BayesFlow 采用无监督学习方式，能够直接从输入数据中学习关于参数的条件概率分布，从而实现对真实后验分布的有效近似(Radev et al., 2022)。BayesFlow 的最大优势在于它能够直接对参数进行估计，而不需要对数据大小和后验分布的形态做出预先假设。此外，它得益于现代 GPU 的加速，能够实现高度并行的计算，这使得 BayesFlow 在认知神经科学领域得到了更多的应用。例如，BayesFlow 已应用于涉及超过百万样本的大规模老龄化研究，能快速地基于大量数据对后验进行估计(von Krause et al., 2022)。BayesFlow 也被用于包含复杂动态参数变化的模型，其中模型参数所代表的潜在认知状态随时间变化，更贴近真实情景，但其参数难以使用传统方法进行推断(Schumacher et al., 2023, 2024)。此外，BayesFlow 在解决复杂模型和高维参数上的优势可用于整合不同模态数据。例如，Ghaderi-Kangavari et al., (2023)结合 BayesFlow 方法对行为数据与神经信号进行联合建模，可用于分析单试次 EEG 测量数据和行为数据。最后，BayesFlow 方法有对应的 Python 软件

包，研究者可快速实施和应用神经后验估计方法(Radev et al., 2022; Radev, Schmitt, Schumacher, et al., 2023)

然而，BayesFlow 类的神经后验估计方法亦有其局限性。这类方法的最大局限在于缺乏灵活性，每当参数有变化时均需要重新训练神经网络。例如，当需要根据实验条件比较参数时，意味着为额外的条件增加参数，而模型估计的参数数量是固定的，因此，增加实验条件意外着需要重新训练一个类似的模型。同样，神经后验估计无法完成具有约束参数的推断需要，如无法根据先前研究的最优参数约束当前模型，导致无法评估数据对特定参数的影响程度。此外，目前的神经后验估计亦无法处理分层模型，仅能针对单个被试或条件进行估计。相较之下，神经似然估计方法在这方面更具优势，无需针对不同先验、实验条件和分层模型进行额外训练。最后，神经后验估计通常依赖于摘要统计网络，这对推断性能至关重要，但在某些实际场景中，构建合适的汇总网络可能存在困难(Radev et al., 2022)。

3.3. 神经模型比较. 模型评估和比较是认知建模中另一个重要的环节（见郭鸣谦等印刷中），但神经似然与神经后验估计均不适用于模型评估和比较，原因在于神经后验估计跳过计算似然值和边际似然，而神经似然估计虽然能够用于计算似然值和边际似然，但处理复杂模型时计算成本高昂(Mancini et al., 2023; Radev et al., 2021)。基于神经似然估计的模型比较的另一个局限在于，需要为每个待比较的模型单独训练神经网络，增加了实现的成本。

针对这些问题，神经模型比较的思路被提出，旨在基于模拟方法识别给定数据最有可能来自于哪个生成模型，从而避免进行参数估计或计算似然和边际似然(Radev et al., 2021)。

神经模型比较方法的核心在于训练一个分类器，用以执行监督学习任务，其中输入为多个生成模型产生的模拟数据，输出则是与这些模型相对应的标签。例如，Radev 等人(2021)提出证据网络(Evidential Network)，该网络接受输入数据$\widetilde{D}_i^{1:M_i}$，并输出模型$\mathcal{M}_j$的后验概率$p\left(\mathcal{M}_j \mid D\right)$，可表示为$f_\phi: D \rightarrow p\left(\mathcal{M}_j \mid D\right)$，其$J$为模型数量。这意味着一旦模型完成训练，就能够根据输入的真实数据，推断出最有可能产生数据的模型。

证据网络的优势在于能够近似贝叶斯因子和模型后验概率，以此估计绝对和相对证据并量化估计的不确定性(Radev et al., 2021)。此外，证据网络采用分摊方法避免了对每个候选模型和数据集单独进行训练的步骤，且无需更改网络结构，仅需调整输出层即可实现对于新模型的评估(Radev et al., 2021)。

证据网络的局限在于，其基于边际似然进行模型比较，因而无法依据后验预测进行模型的评估和比较（见郭鸣谦等 印刷中）。此外，证据网络不适用于处理具有高维参数空间的模型，尤其是对于贝叶斯分层模型而言。针对于此，Elsemüller 等(2023)提出深度分层模型比较(deep hierarchical model comparision, DHMC)方法，旨在比较具有多层结构和高维参数的贝叶斯模型。DHMC 方法在证据网络的基础上使用两组不变性网络分别处理试次和被试数量不一致的问题，在多种贝叶斯模型比较场景中展现出与桥接抽样技术相当或更优的性能，适用于广泛的模型比较任务。

4. 总结和展望

认知建模作为模拟和理解复杂心理过程的工具，正日益获得重视。一方面，计算模型可以整合多模态数据，量化个体差异，并能为精神疾病提供诊断依据(Eaton et al., 2023; Kriegeskorte & Douglas, 2018; Wang & Krystal, 2014)。另一方面，生成模型的逆问题、模型分层化为认知建模的进一步应用带来严峻考验(Turner & Zandt, 2018)。基于模拟的推断方法被开发以解决上述挑战，为认知模型的发展提供了新的途径。由于在处理高维数据和复杂模型方面的显著优势，神经网络被纳入模拟推断方法，并在近似似然、后验估计和模型比较中初步显示出应用价值。

神经模拟推断的核心特征在于其利用了神经网络的分摊特性，将生成模拟数据的成本主要集中在训练阶段，一旦训练完成，神经网络模型便能实现快速推理(Radev et al., 2022)。此外，神经模拟推断基于神经条件密度估计方法，能够对似然函数、参数后验和似然比分布的概率密度进行有效估计(Papamakarios, 2019)。随着标准化流技术的进步，神经模拟推断的性能和效率得到显著提升(Durkan et al., 2019; Papamakarios et al., 2021)。相较于传统的模拟推断方法，神经模拟推断有效优化了计算资源的使用，并利用神经网络处理高维数据的强大能力，克服了维度诅咒问题，减少了对摘要统计的依赖。但是，神经模拟推断也存在一些待解决的问题。

并非所有研究者都认同分摊策略，相反，一些研究者认为分摊策略实际上是一种对于模拟数据的浪费(Greenberg et al., 2019; Papamakarios & Murray, 2016)。例如，与顺序神经密度估计相比，分摊神经网络通常需要更多的模拟来充分探索参数空间。同样，在近似贝叶斯计算和概率密度近似方法中，对于相对简单的模型和能够迅速生成模拟数据的场景，投入大量时间训练神经网络可能并不划算。分摊方法的优势在于一旦训练完成，即可应用于大量数据。Radev 等(2022)提出了模拟预算收支平衡的概念，用以量化在处理多大规模的数据时，非分摊方法的推理时间会超过分摊方法的训练时间(Radev et al., 2022; Radev,

Schmitt, Pratz, et al., 2023)。例如，在完成包含超过 75 份数据集的推理任务时，分摊方法展现出相对于其他方法的优势(Radev et al., 2022)。

4.1. 神经网络训练花费的权衡. 因此，在处理包含数千人的大型数据集时，例如应用于心理测量模型或辅助医疗诊断模型，分摊方法具有显著的优势。此外，在比较多个结构相似的认知模型时，需要分别将每个模型重复拟合多个数据集，在此情况下，采用分摊方法更为有效。最后，先验范围对分摊方法至关重要。若先验范围无法事先确定，可能需要选择更广泛的先验范围，这虽然增强了神经网络模型的泛化能力，但同时也会增加训练的难度。因此，如果能够预先确定一个较窄的先验范围，或者专注于探索参数空间的一个特定区域而非追求广泛的泛化能力，那么采用顺序神经网络可能是一个更合适的选择。

4.2. 三种神经模拟推断方法的整合. 尽管神经模拟推断存在三种应用方向，但三者彼此各司其职、互不相通。近期，研究者们开始尝试将这些应用方向整合在一起。例如，Radev 等(2023)提出名为 JANA(Jointly Amortized Neural Approximation)的整合框架，旨在同时实现对似然和后验神经网络的训练，其核心在于采用条件标准流技术同时限制对似然和后验密度的估计。通过这种方式，JANA 能够通过神经后验网络进行参数估计，并利用参数样本结合神经似然网络计算模型比较指标。实证研究表明，JANA 能够获得与传统方法相媲美的 LOO-CV 和边际似然值的估计效果(Radev, Schmitt, Pratz, et al., 2023)。Mancini 等人(2023)最近提出了一个更为全面的框架，该框架不仅支持将神经后验网络与神经似然网络相结合以进行参数估计和模型比较，而且还允许通过融合神经似然比网络与神经似然网络来实现相同的目标。

4.3. 神经网络训练的有效性. 通常情况下，采用成熟单一的推断方法已足够解决特定的认知建模问题。例如，利用神经后验估计方法可以快速估计模型参数，进而利用参数进行后续统计分析，适用于大规模数据；利用神经似然网络可以替代不存在解析似然或解析似然难以计算的模型，以此结合 MCMC 等方法进行参数估计，以及进一步通过似然值计算模型评估指标，适用于具有相同似然的认知模型。但要进行全面的模型评估，建议采用联合方法，以实现更综合、精确和高效的推断结果。例如，利用神经后验估计或神经似然比估计进行参数估计，接着使用神经似然网络计算似然值；或是利用神经后验估计的结构反过来验证神经似然网络和神经似然比网络在 MCMC 下的稳定性。

神经模拟推断的另一个关键问题在于如何确保所训练神经网络的有效性。神经模拟推断的表现主要依赖于生成模型的准确性。不准确的生成模型设定会对神经网络的推断性能造成严重影响，并且尚无万全之策能规避所有潜在的错误(Cannon et al., 2022)。Hermans 等(2021)指出，目前流行的神经模拟推断算法，包括神经后验估计、神经似然度估计以及各

类近似贝叶斯计算方法，均有可能生成过度自信的后验近似结果。为增强估计的稳健性，他们建议集成多个神经网络，通过平均这些网络的估计来实现更可靠的推断。

然而，考虑到设计一个好的生成模型本身需要丰富的专业知识和经验，神经网络的不良表现可能揭示了生成模型本身的缺陷，促使研究者重新审视其生成模型的合理性(Kelly et al., 2023)。例如，某些模型参数可能表现出“滑坡”和“膨胀”的特性，即模型预测对参数变化不敏感(Holmes & Trueblood, 2018)。这种特性既可被视为模型具有鲁棒性和泛化能力，也可能意味着模型表现不佳，如无法实现参数的精确恢复。因此，关键在于如何有效地评估模型，这包括使用分布外验证、模型比较、后验预测检验和模拟研究等方法。最近，研究者提出最大平均差异(Maximum Mean Discrepancy, MMD)，用于识别模型错误(Schmitt et al., 2024)。该方法能够检测先验范围对数据预测与真实数据之间差异的影响，并可用于优化神经网络训练过程。

简而言之，在具备专业知识的情况下，设计生成模型和优化神经网络应相辅相成：利用神经网络快速推理的能力，高效地对生成模型进行评估，识别其不足，并据此进行优化，进而推动认知理论和模型的进步。

4.4. 展望

神经模拟推断技术在认知建模中具有巨大潜力，是研究者进行模型的扩展和理论创新的有力工具。例如，在认知科学与神经科学融合的背景下，Ghaderi-Kangavari 等人(2023) 基于神经模拟推断技术开发了新的联合建模框架，该建模框架通过共享的潜变量实现对知觉行为和脑电信号特征的同步预测。该联合模型可扩展至多种决策任务并与多种神经成像数据进行融合，具有较大的潜力。此外，神经模拟推断技术有望加速认知模型在测量模型方面的应用，例如在大型老龄化研究项目中的应用(von Krause et al., 2022)，以及在计算精神病学领域的诊断应用(Eaton et al., 2023; Ratcliff & McKoon, 2022)。

当前，神经模拟推断主要应用于证据积累模型(Ghaderi-Kangavari et al., 2023; Wieschen et al., 2023)，尽管也有些研究涉及强化学习模型和多项式决策树模型(Elsemüller et al., 2023; Rmus et al., 2024)，但未来研究应更多探索其他模型。此外，心理学数据的一个重要特征是分层结构，但目前神经模拟推断主要关注分层模型的比较(Elsemüller et al., 2023)，缺乏关于分层模型参数估计方面的研究。最后，尽管已有大量研究关注神经模拟推断在模型评估和比较方面的应用，但这些方法还无法完全利用贝叶斯模型平均的思想来整合不同模型的估计。

参考文献

郭鸣谦, 潘晚坷, 胡传鹏. (印刷中). *认知建模中模型比较的方法*. 心理科学进展.

Boehm, U., Evans, N. J., Gronau, Q. F., Matzke, D., Wagenmakers, E.-J., & Heathcote, A. J. (2023). Inclusion Bayes factors for mixed hierarchical diffusion decision models. *Psychological Methods*. https://doi.org/10.1037/met0000582

Boelts, J., Lueckmann, J.-M., Gao, R., & Macke, J. H. (2022). Flexible and efficient simulation-based inference for models of decision-making. *eLife*, *11*, e77220. https://doi.org/10.7554/eLife.77220

Brehmer, J., Louppe, G., Pavez, J., & Cranmer, K. (2020). Mining gold from implicit models to improve likelihood-free inference. *Proceedings of the National Academy of Sciences*, *117*(10), 5242–5249. https://doi.org/10.1073/pnas.1915980117

Cannon, P., Ward, D., & Schmon, S. M. (2022). *Investigating the Impact of Model Misspecification in Neural Simulation-based Inference* (No. arXiv:2209.01845). arXiv. http://arxiv.org/abs/2209.01845

Churchland, P. S., & Sejnowski, T. J. (1988). Perspectives on cognitive neuroscience. *Science*, *242*(4879), 741–745. https://doi.org/10.1126/science.3055294

Cranmer, K., Brehmer, J., & Louppe, G. (2020). The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, *117*(48), 30055–30062. https://doi.org/10.1073/pnas.1912789117

Durkan, C., Bekasov, A., Murray, I., & Papamakarios, G. (2019). Neural spline flows. *Advances in Neural Information Processing Systems*, *32*. https://proceedings.neurips.cc/paper/2019/hash/7ac71d433f282034e088473244df8c02-Abstract.html

Eaton, N. R., Bringmann, L. F., Elmer, T., Fried, E. I., Forbes, M. K., Greene, A. L., Krueger, R. F., Kotov, R., McGorry, P. D., Mei, C., & Waszczuk, M. A. (2023). A review of approaches and models in psychopathology conceptualization research. *Nature Reviews Psychology*, *2*(10), Article 10. https://doi.org/10.1038/s44159-023-00218-4

Elsemüller, L., Schnuerch, M., Bürkner, P.-C., & Radev, S. (2023). *A Deep Learning Method for Comparing Bayesian Hierarchical Models*. https://doi.org/10.48550/arXiv.2301.11873

Farrell, S., & Lewandowsky, S. (2018). *Computational modeling of cognition and behavior*. Cambridge University Press. https://doi.org/10.1017/CBO9781316272503

Fengler, A., Govindarajan, L. N., Chen, T., & Frank, M. J. (2021). Likelihood approximation networks (LANs) for fast inference of simulation models in cognitive neuroscience. *eLife*, *10*, e65074. https://doi.org/10.7554/eLife.65074

Forstmann, B. U., & Turner, B. M. (Eds.). (2024). *An introduction to model-based cognitive neuroscience*. Springer International Publishing. https://doi.org/10.1007/978-3-031-45271-0

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, *24*(6), 997–1016. https://doi.org/10.1007/s11222-013-9416-2

Geng, H., Chen, J., Chuan-Peng, H., Jin, J., Chan, R. C. K., Li, Y., Hu, X., Zhang, R.-Y., & Zhang, L. (2022). Promoting computational psychiatry in China. *Nature Human Behaviour*, *6*(5), 615–617. https://doi.org/10.1038/s41562-022-01328-4

Ghaderi-Kangavari, A., Rad, J. A., & Nunez, M. D. (2023). A General Integrative Neurocognitive Modeling Framework to Jointly Describe EEG and Decision-making on Single Trials. *Computational Brain & Behavior*, *6*(3), 317–376. https://doi.org/10.1007/s42113-023-00167-4

Gonçalves, P. J., Lueckmann, J.-M., Deistler, M., Nonnenmacher, M., Öcal, K., Bassetto, G., Chintaluri, C., Podlaski, W. F., Haddad, S. A., Vogels, T. P., Greenberg, D. S., & Macke, J. H. (2020). Training deep neural density estimators to identify mechanistic models of neural dynamics. *eLife*, *9*, e56261. https://doi.org/10.7554/eLife.56261

Greenberg, D. S., Nonnenmacher, M., & Macke, J. (2019, May 17). *Automatic posterior transformation for likelihood-free inference*. International Conference on Machine Learning. https://www.semanticscholar.org/paper/Automatic-Posterior-Transformation-for-Inference-Greenberg-Nonnenmacher/9d5233ea2135e7bc9023127af5b11a6af92e19cd

Gronau, Q. F., Sarafoglou, A., Matzke, D., Ly, A., Boehm, U., Marsman, M., Leslie, D. S., Forster, J. J., Wagenmakers, E.-J., & Steingroever, H. (2017). A tutorial on bridge sampling. *Journal of Mathematical Psychology*, *81*, 80–97. https://doi.org/10.1016/j.jmp.2017.09.005

Guest, O., & Martin, A. E. (2021). How Computational Modeling Can Force Theory Building in Psychological Science. *Perspectives on Psychological Science*, 174569162097058–174569162097058. https://doi.org/10.1177/1745691620970585

Hashemi, M., Vattikonda, A. N., Jha, J., Sip, V., Woodman, M. M., Bartolomei, F., & Jirsa, V. K. (2023). Amortized bayesian inference on generative dynamical network models of epilepsy using deep neural density estimators. *Neural Networks*, *163*, 178–194. https://doi.org/10.1016/j.neunet.2023.03.040

Hermans, J., Delaunoy, A., Rozet, F., Wehenkel, A., Begy, V., & Louppe, G. (2021, October 13). *A trust crisis In simulation-based inference? Your posterior approximations can Be unfaithful*. https://www.semanticscholar.org/paper/A-Trust-Crisis-In-Simulation-Based-Inference-Your-Hermans-Delaunoy/608b513bc049318248905216830ac292aef5de9e

Holmes, W. R. (2015). A practical guide to the probability density approximation (PDA) with improved implementation and error characterization. *Journal of Mathematical Psychology*, *68–69*, 13–24. https://doi.org/10.1016/j.jmp.2015.08.006

Holmes, W. R., & Trueblood, J. S. (2018). Bayesian analysis of the piecewise diffusion decision model. *Behavior Research Methods*, *50*(2), 730–743. https://doi.org/10.3758/s13428-017-0901-y

Kangasrääsiö, A., Jokinen, J. P. P., Oulasvirta, A., Howes, A., & Kaski, S. (2019). Parameter inference for computational cognitive models with approximate bayesian computation. *Cognitive Science*, *43*(6), e12738. https://doi.org/10.1111/cogs.12738

Kelly, R. P., Nott, D. J., Frazier, D. T., Warne, D. J., & Drovandi, C. (2023). *Misspecification-robust sequential neural likelihood for simulation-based inference*. https://doi.org/10.48550/ARXIV.2301.13368

Kriegeskorte, N., & Douglas, P. K. (2018). Cognitive computational neuroscience. *Nature Neuroscience*, *21*(9), 1148–1160. https://doi.org/10.1038/s41593-018-0210-5

Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Academic Press.

Liu, Q., Xu, J., Jiang, R., & Wong, W. H. (2021). Density estimation using deep generative neural networks. *Proceedings of the National Academy of Sciences*, *118*(15), e2101344118. https://doi.org/10.1073/pnas.2101344118

Ma, W. J. (2022). *Bayesian Models of Perception and Action*.

Mancini, A. S., Docherty, M. M., Price, M. A., & McEwen, J. D. (2023). Bayesian model comparison for simulation-based inference. *RAS Techniques and Instruments*, *2*(1), 710–722. https://doi.org/10.1093/rasti/rzad051

McClelland, J. L. (2009). The place of modeling in cognitive science. *Topics in Cognitive Science*, *1*(1), 11–38. https://doi.org/10.1111/j.1756-8765.2008.01003.x

Palestro, J. J., Sederberg, P. B., Osth, A. F., Van Zandt, T., & Turner, B. M. (2018). *Likelihood-free methods for cognitive science*. https://doi.org/10.1007/978-3-319-72425-6

Papamakarios, G. (2019). *Neural density estimation and likelihood-free inference* (No. arXiv:1910.13233). arXiv. http://arxiv.org/abs/1910.13233

Papamakarios, G., & Murray, I. (2016). *Fast ε-free inference of simulation models with bayesian conditional density estimation*. Neural Information Processing Systems. https://www.semanticscholar.org/paper/Fast-%CE%B5-free-Inference-of-Simulation-Models-with-Papamakarios-Murray/5202a43d22de63139ee2bb6802bedc738dfc9235

Papamakarios, G., Nalisnick, E., Rezende, D. J., Mohamed, S., & Lakshminarayanan, B. (2021). Normalizing flows for probabilistic modeling and inference. *Journal of Machine Learning Research*, *22*(57), 1–64. https://doi.org/10.48550/arXiv.1912.02762

Papamakarios, G., Pavlakou, T., & Murray, I. (2017). Masked autoregressive flow for density estimation. *Advances in Neural Information Processing Systems*, *30*. https://proceedings.neurips.cc/paper_files/paper/2017/hash/6c1da886822c67822bcf3679d04369fa-Abstract.html

Radev, S. T., D'Alessandro, M., Mertens, U. K., Voss, A., Kothe, U., & Burkner, P.-C. (2021). Amortized Bayesian Model Comparison With Evidential Deep Learning. *IEEE Transactions on Neural Networks and Learning Systems*, 1–15. https://doi.org/10.1109/TNNLS.2021.3124052

Radev, S. T., Mertens, U. K., Voss, A., Ardizzone, L., & Kothe, U. (2022). BayesFlow: Learning Complex Stochastic Models With Invertible Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(4), 1452–1466. https://doi.org/10.1109/TNNLS.2020.3042395

Radev, S. T., Mertens, U. K., Voss, A., & Köthe, U. (2020). Towards end-to-end likelihood-free inference with convolutional neural networks. *British Journal of Mathematical and Statistical Psychology*, *73*(1), 23–43. https://doi.org/10.1111/bmsp.12159

Radev, S. T., Schmitt, M., Pratz, V., Picchini, U., Köthe, U., & Bürkner, P.-C. (2023). *JANA: Jointly Amortized Neural Approximation of Complex Bayesian Models*.

Radev, S. T., Schmitt, M., Schumacher, L., Elsemüller, L., Pratz, V., Schälte, Y., Köthe, U., & Bürkner, P.-C. (2023). BayesFlow: Amortized bayesian workflows with neural networks. *Journal of Open Source Software*, *8*(89), 5702. https://doi.org/10.21105/joss.05702

Radev, S. T., Voss, A., Wieschen, E. M., & Burkner, P.-C. (2020). Amortized bayesian inference for models of cognition. *ArXiv*. https://www.semanticscholar.org/paper/Amortized-Bayesian-Inference-for-Models-of-Radev-Voss/2c719c908bf8e27883ae80f0c42453e33dcf110a

Ratcliff, R., & McKoon, G. (2022). Can neuropsychological testing be improved with model-based approaches? *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2022.08.015

Raynal, L., Marin, J.-M., Pudlo, P., Ribatet, M., Robert, C. P., & Estoup, A. (2019). ABC random forests for bayesian parameter inference. *Bioinformatics*, *35*(10), 1720–1728. https://doi.org/10.1093/bioinformatics/bty867

Rmus, M., Pan, T.-F., Xia, L., & Collins, A. G. E. (2024). Artificial neural networks for model identification and parameter estimation in computational cognitive models. *PLoS Computational Biology*, *20*(5), e1012119. https://doi.org/10.1371/journal.pcbi.1012119

Robert, C. P., Cornuet, J.-M., Marin, J.-M., & Pillai, N. S. (2011). Lack of confidence in approximate bayesian computation model choice. *Proceedings of the National Academy of Sciences*, *108*(37), 15112–15117. https://doi.org/10.1073/pnas.1102900108

Schad, D. J., Betancourt, M., & Vasishth, S. (2021). Toward a principled Bayesian workflow in cognitive science. *Psychological Methods*, *26*(1), 103–126. https://doi.org/10.1037/met0000275

Schmitt, M., Bürkner, P.-C., Köthe, U., & Radev, S. T. (2024). *Detecting model misspecification in amortized bayesian inference with neural networks* (U. Köthe & C. Rother, Eds.; Vol. 14264, pp. 541–557). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-54605-1_35

Schumacher, L., Bürkner, P.-C., Voss, A., Köthe, U., & Radev, S. T. (2023). Neural superstatistics for bayesian estimation of dynamic cognitive models. *Scientific Reports*, *13*(1), 13778. https://doi.org/10.1038/s41598-023-40278-3

Schumacher, L., Schnuerch, M., Voss, A., & Radev, S. T. (2024). *Validation and comparison of non-stationary cognitive models: A diffusion model application* (No. arXiv:2401.08626). arXiv. http://arxiv.org/abs/2401.08626

Tejero-Cantero, A., Boelts, J., Deistler, M., Lueckmann, J.-M., Durkan, C., Gonçalves, P. J., Greenberg, D. S., & Macke, J. H. (2020). *SBI -- a toolkit for simulation-based inference* (No. arXiv:2007.09114). arXiv. https://doi.org/10.48550/arXiv.2007.09114

Turner, B. M., & Sederberg, P. B. (2014). A generalized, likelihood-free method for posterior estimation. *Psychonomic Bulletin & Review*, *21*(2), 227–250. https://doi.org/10.3758/s13423-013-0530-0

Turner, B. M., & Zandt, T. V. (2018). Approximating Bayesian Inference through Model Simulation. *Trends in Cognitive Sciences*, *22*(9), 826–840. https://doi.org/10.1016/j.tics.2018.06.003

Vandekerckhove, J., Tuerlinckx, F., & Lee, M. D. (2011). Hierarchical Diffusion Models for Two-Choice Response Times. *Psychological Methods*, *16*(1), 44–62. https://doi.org/10.1037/a0021765

von Krause, M., Radev, S. T., & Voss, A. (2022). Mental speed is high until age 60 as revealed by analysis of over a million participants. *Nature Human Behaviour*, *6*(5), 700–708. https://doi.org/10.1038/s41562-021-01282-7

Wang, X.-J., & Krystal, J. H. (2014). Computational psychiatry. *Neuron*, *84*(3), 638–654. https://doi.org/10.1016/j.neuron.2014.10.018

Wieschen, E. M., Makani, A., Radev, S. T., Voss, A., & Spaniol, J. (2023). Age-related differences in decision-making: Evidence accumulation is more gradual in older age. *Experimental Aging Research*. https://www.tandfonline.com/doi/abs/10.1080/0361073X.2023.2241333

Wilson, R. C., & Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *eLife, 8,* e49547. https://doi.org/10.7554/eLife.49547

附录

贝叶斯认知模型的数学表示方式

在贝叶斯框架下，认知模型的模拟器或生成形式可以表示为:

$$D \sim p(D|\theta) \tag{1}$$

θ为模型参数，代表潜在认知状态。D为生成数据，代表在相应认知状态下所产生的行为或观测数据。$p(D|\theta)$为条件概率分布，表示在给定参数θ下观察到的数据D的概率，其暗含了认知加工的机制，即潜在认知状态对于加工过程和行为的影响。

从推论的角度来看，研究者关注的是如何从实验数据中推断潜在的心理过程，即进行参数估计。与传统的频率参数推断方法（如最小二乘法）相比，贝叶斯统计的吸引力在于它允许同时评估模型参数的估计值及其不确定性，而不仅仅是提供精确的点估计。参数的后验分布可以通过贝叶斯公式表示：

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)} \tag{2}$$

$p(\theta|D)$为参数θ的后验分布，表示在获取数据D后对于参数θ的推断，其概率分布的形式蕴含了参数及其所代表的心理状态的不确定性。后验分布依赖于两个关键函数的准确评估：模型参数的先验分布$p(\theta)$，代表在未获得数据前对于参数分布的信念；以及似然函数$p(D|\theta)$，也可以写作$L(D|\theta)$[1]，代表了在已知参数的条件下数据出现的概率。然而，对于具有复杂认知机制的模型，受到生成模型逆问题所带来的困扰，确定其似然函数通常更加困难。最后，$p(D)$为贝叶斯证据(evidence)，也称为边际似然(marginal likelihood)，表示在所有参数范围内观察到数据D的概率的总和，数学上表示为$p(D) = \int p(D|\theta)p(\theta)d\theta$。

贝叶斯框架下的模型比较 在贝叶斯模型比较框架下，多个竞争生成模型被表示为集合$\mathcal{M} = \{\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_j\}$，每个模型中参数的后验分布表示为，其中$J = 1{:}j$为模型序号：

[1] 可以发现，$p(D|\theta)$既可以表示生成模型，也可以表示似然函数。这是因为对于简单的概率分布而言，如正态分布，从分布中生成数据D，和反过来根据数据D计算似然值均相对容易实现。因此，为了避免概念上的混淆，我们将在后续的讨论中使用$L(D|\theta)$表示似然函数，而$p(d|\theta)$表示生成函数。注意，在一些统计和机器学习领域习惯将似然函数写作$L(\theta|D)$，而非$L(D|\theta)$。

$$p(\theta_j|D,\mathcal{M}_j)=\frac{p(D|\theta_j,\mathcal{M}_j)p(\theta|\mathcal{M}_j)}{p(D|\mathcal{M}_j)} \tag{3}$$

与公式2)相比，贝叶斯推断中的所有元素均受到模型的条件概率的影响。其中，边际似然(marginal likelihood)或贝叶斯证据(Bayesian evidence)为$p(D\mid\mathcal{M}_j)=\int p(D\mid\theta_j,\mathcal{M}_j)p(\theta_j\mid\mathcal{M}_j)\mathrm{d}\theta_j$，表示模型$\mathcal{M}_j$生成数据$D$的概率。边际似然作为模型选择的核心，为计算贝叶斯因子和模型后验概率奠定了基础。其中，贝叶斯因子通过边际似然的比率来量化模型间的相对支持度，定义为 $\mathrm{BF}_{jk}=\frac{p(x|\mathcal{M}_j)}{p(x|\mathcal{M}_k)}$，其中$\mathcal{M}_j,\mathcal{M}_k$分别代表两个竞争模型。当$\mathrm{BF}_{jk}>1$，表示在给定数据$D$的情况下，模型$\mathcal{M}_j$比模型$\mathcal{M}_k$更受青睐(胡传鹏 等, 2018)。另一方面，结合边际似然和模型先验$p(\mathcal{M}_j)$可以计算得到模型后验概率，以直接量化在指定数据D下模型$\mathcal{M}_j$出现的概率，其数学形式为：

$$p(\mathcal{M}_j\mid D)=\frac{p(D\mid\mathcal{M}_j)p(\mathcal{M}_j)}{\sum_{j=1}^{J}p(D\mid\mathcal{M}_j)p(\mathcal{M}_j)} \tag{4}$$

贝叶斯因子和模型后验概率的计算均依赖于边际似然度，但由于边际似然度通常涉及高维积分，这导致其直接计算变得复杂或不可行。证据网络的核心思路在于利用神经网络直接从数据中估计模型后验概率，从而巧妙地规避了边际似然的计算难题。

传统模拟推断推断方法数学表达形式

在认知建模领域，模拟推断技术亦被广泛采用和研究，主要表现为两种方法：近似贝叶斯计算和概率密度近似。

近似贝叶斯计算 近似贝叶斯计算(Approximate Bayesian Computation，ABC)是一个算法系列，最基础的实现是基于拒绝的近似贝叶斯计算(rejection-based ABC)。其核心思想在于提供一种替代或绕过计算似然函数$L(\theta|D)$的方案，通过模拟数据的生成和处理，间接逼近难以直接计算的后验分布$p(\theta|D)$。该方法包括 5 个核心步骤(Palestro et al., 2018)：首先从先验分布$\pi(\theta)$中抽取参数θ^*的样本；进而利用参数生成模拟数据$\widetilde{D}\sim p(D|\theta^*)$(见公式1))；同时计算模拟数据数据和真实数据的摘要统计量(summary statistic)，分别得到$S(\widetilde{D})$和$S(D)$；随后比较模拟数据与观测数据的摘要统计之间距离(接近程度)，$d=\rho\left(S(\widetilde{D}),S(D)\right)$[2]；最后，依据预设的距离度量$\epsilon$评估摘要统计距离$d$，如果$d<\epsilon$，则接

[2] 在近似贝叶斯计算方法中，计算摘要统计量距离的函数旨在衡量模拟数据摘要统计量与实际数据摘要统计量之间的差异，包括欧几里得距离、曼哈顿距离、切比雪夫距离、马氏距离、夹角余弦以及基于核的距离等(Bhattacharya et al., 2007)。

受参数为后验样本，即$\theta_i = \theta^*$。重复整个流程直至累积到足够的后验样本量，进而构建出后验分布的近似。当 $\epsilon \to 0$ 时，该方法推理的准确性会得到提高，但会使得接受概率降低，进而导致需要进行更多次数据模拟。反过来说，较大的 ϵ 值会提高获得后验样本效率，但代价是牺牲推理的质量(Cranmer et al., 2020)。

概率密度近似 鉴于摘要统计需要手动设置距离度量，部分研究者开始采用概率密度近似(Probablistic Density Approximation, PDA)以替代传统的摘要统计，旨在实现更为准确的参数估计(Holmes, 2015; Turner & Sederberg, 2014)。

概率密度近似是一种非参数估计方法，其核心思想在于运用整个数据的分布信息而非仅靠汇总统计量来形成对于似然函数的近似，称为伪似然(pseudo-likelihood)，记作$\hat{L}(\theta^*|\tilde{D})$。具体来说，该方法需要先根据参数生成一组模拟数据，$\tilde{D} = \{\tilde{d}_\iota\}$，其中$i$取 1 到$N$，表示$N$个模拟数据。之后，利用该数据集构建概率密度函数，作为似然函数的近似，数学形式为：

$$L(\theta|d) \approx \hat{L}(\theta|d) \coloneqq \frac{1}{N}\sum_{i=1}^{N} K_h\left(d - \tilde{d}_\iota\right) \tag{5}$$

其中，d表示待估计似然值的真实数据，可以为单个数据点或多个数据；$K_h\left(d - \tilde{d}_\iota\right)$表示核密度估计(Kernel Density Estimation，KDE)方法。核密度估计是一种平滑技术，通过加权平均相邻数据点来估计未知观测数据的概率密度函数，这使得它能够基于现有数据集推断出新数据点的概率分布情况，从而用于密度估计(Turner & Sederberg, 2014)。

由于概率密度近似能近似似然函数的特性，使其展现出广泛的应用潜力。该方法不仅能直接用于频率优化方法，还能与 MCMC 方法结合。此外，概率密度近似所计算的近似似然值可进一步用于计算模型评估指标，使其在模型比较方面相对于近似贝叶斯计算方法具有一定的优势，这一议题已有相关讨论(Holmes, 2015; Robert et al., 2011)。

神经网络相关核心技术

分摊(Amortization) 神经密度估计技术能够用于估计后验概率分布和似然函数，其特征在于分摊(Amortization)特性，其目标在于将计算资源集中用于训练一个单一模型，该模型能够一次性处理多个数据点，并且具有泛化到新数据的能力(见图 1)。分摊网络的优势在于一旦模型训练完成，可以对新的数据点进行快速推断，而无需进行额外训练(Radev et al., 2022; Radev, Schmitt, Pratz, et al., 2023)。

以神经后验估计为例，分摊策略包括三个主要步骤：在模拟阶段，生成大量训练数据对$\{\tilde{\theta}_i, \tilde{D}_i^{1:M}\}_{i=1}^{N}$，其中$i$取 1 到$N$，表示$N$份数据对，$\tilde{D}_i$由 1 到$M$个数据点构成；在训练阶

段，调整神经网络权重ϕ使得神经网络q_{ϕ}学习数据D和概率密度$p(\theta|D)$的映射；在推理阶段，输入真实观测数据D，运行神经网络后得到参数后验样本$p(\theta|D) := q_{\phi}(\theta|D)$。分摊策略的关键在于训练和推理完全分离，一旦网络训练完成，神经网络可以离线运行，以极快的速度获得推理，即训练过程分摊了推理的高额运算(Radev, Voss, et al., 2020)。此外，由于模型是基于整个先验预测分布产生的数据集上训练的，因此它通常具有更好的泛化能力，能够更好地处理新数据(Radev, Voss, et al., 2020)。

标准化流(Normalizing Flow, NF) 标准化流(Normalizing Flow, NF, Papamakarios, 2019; Papamakarios et al., 2021)采用生成模型架构，可实现对条件密度的直接估计，并且其所需的训练数据更少。

标准化流作为一种可逆生成模型，执行无监督学习任务，无需借助核密度估计计算似然值作为训练标签。其核心思想在于通过连续的、可逆的变换将一个简单的分布$z \sim \pi(z)$(如高斯分布)转化为更复杂的数据分布。

在估计似然函数时，标准化流通过条件可逆神经网络(conditional invertible neural network, cINN)实现对条件概率密度的估计(Durkan et al., 2019; Papamakarios et al., 2021)，其数学形式为：

$$L_{\phi}(D \mid \theta) = \pi(z_D)\left|\det\left(\frac{\partial f^{-1}}{\partial D}\right)\right| \tag{6}$$

标准化流中，似然函数$L_{\phi}(D \mid \theta)$由神经网络参数化，其权重为ϕ，通过一个简单分布$\pi(z_D)$变换得到。其中，f^{-1}为可逆变换，可以将数据映射到简单分布的潜在空间，即$z_D = f(D)$，反之亦然，$D = f^{-1}(z_D)$。$\left|\det\left(\frac{\partial f^{-1}}{\partial D}\right)\right|$为雅可比矩阵的行列式，为可逆变换的局部缩放因子，衡量数据D到z_D映射过程的缩放性。正如其名，“标准化流”由一系列的可逆神经网络模块串接形成，其作用是首先将数据标准化为潜变量，然后通过一系列变换(即“流”)将其转化为目标分布(Papamakarios et al., 2021)。因此，标准化流的关键在于定义和学习可逆变换。

在估计参数后验分布$p(\theta \mid D)$时，可以表示为在给定数据D的条件下参数θ和潜在变量z_{θ}(来自于简单分布)之间的标准化流：

$$p_{\phi}\left(\theta \,\middle|\, \mathcal{H}_{\eta}(D)\right) = p(z_{\theta})\left|\det\left(\frac{\partial f^{-1}}{\partial \theta}\right)\right| \tag{7}$$

其中，f^{-1}为可逆变换；雅各比矩阵的行列式$\left|\det\left(\frac{\partial f^{-1}}{\partial \theta}\right)\right|$衡量了简单分布到后验分布的缩放性。

标准化流已发展出多种高效的可逆变换方法，包括仿射(affine)，多项式样条(polynomial spline)变化，和立方样条(cubic spline)变换(Durkan et al., 2019; Papamakarios et al., 2021)。MNLE 采用了神经样条流(neural spline flow, NSF)的架构，通过几个单调有理二次样条(monotonic rational-quadratic splines)模块实现对似然密度的估计。例如，对于同时包括反应时和反应，该模型可以采用包含两个网络的混合神经网络架构，其中一个网络负责离散变量，针对行为决策反应等离散数据，另一个网络负责连续变量，针对反应时等连续测量数据。

LANs 训练特点 Fengler 等人(2021)提出“似然近似网络”(Likelihood Approximate Networks, LANs)，以快速和准确地估计认知模型的似然函数。LANs 采用多层感知机(Multilayer Perceptrons, MLPs)来学习参数和数据到对数似然函数的映射$f_{\mathrm{MLP}(\phi)}: (\theta, D) \rightarrow \log L\,(D|\theta)$，其中$f_{\mathrm{MLP}(\phi)}$代表具有权重参数$\phi$的多层感知机神经网络。LANs 的训练过程包含两个关键步骤：数据准备和训练。

在数据准备阶段，首先采集一组参数 $\tilde{\theta}_i$；接着根据参数生成模拟数据$\tilde{D}_i^{1:M}$；随后基于核密度估计方法(见概率密度近似小节)计算每个生成数据点的对数似然函数值$\log L\left(\tilde{D}\middle|\tilde{\theta}\right)_i^{1:M}$；重复以上步骤 N 次得到训练数据集$\left\{\tilde{\theta}_i, \tilde{D}_i^{1:M}, \log L\left(\tilde{D}\middle|\tilde{\theta}\right)_i^{1:M}\right\}_{i=1}^{N}$。为确保网络能够有效学习数据的多样化概率分布，训练数据集中的参数从匀分布中抽取，包含 150 万个参数样本，每个样本再输入模拟器生成 10 万个数据点(Fengler et al., 2021)。因此，训练集共包含 1500 亿个带标签的样本以供网络训练。

在模型训练阶段，LANs 执行监督学习任务，其中输入为参数$\tilde{\theta}$和数据$\tilde{D}$，输出为对数似然函数的估计$\log \tilde{L}\left(\tilde{D}\middle|\tilde{\theta}\right)$。通过优化网络权重$\phi$以最小化对数似然估计$\log \tilde{L}\left(\tilde{D}\middle|\tilde{\theta}\right)$与实际对数似然值$\log L\left(\tilde{D}\middle|\tilde{\theta}\right)_i^{1:M}$之间的差异，通过 Huber 损失函数衡量。训练通过随机梯度下降的反向传播算法进行，并采用 Adam 优化算法来调整网络权重(Kingma & Ba, 2014)。

MNLE 训练特点 为了克服 LANs 的局限性，Boelts 等(2022)提出了混合神经似然估计(Mixed Neural Likelihood Estimation, MNLE)。MNLE 采纳标准化流架构，可实现对条件密度的直接估计。相较于 LANs 使用类似“回归”的方式建立数据和参数到似然值的映射，MNLE 可以直接对于似然函数进行估计，并且其所需的训练数据更少(Boelts et al., 2022)。

MNLE 同样涉及两个训练阶段。在数据准备阶段，MNLE 需要构造数据对集$\{\tilde{\theta}_i, \tilde{D}_i\}_{i=1}^{N}$。相较于 LANs，MNLE 具有两点优势：首先，MNLE 不再依赖 KDE 生成似然值作为标签；其次，MNLE 中的参数和数据点之间是一一对应的关系，而无需根据每组参数

生成大量数据点$\widetilde{D}_i^{1:M}$。在训练阶段，MNLE 执行无监督学习任务，通常采用 Adam 优化器和最大似然损失函数(maximum likelihood loss)，即$\sum_{i=1}^{N} \log L_\phi\,(D_i|\theta_i)$。

MNLE 的优势在于其模拟效率。相较于 LANs 需要百亿级别的模拟数据，MNLE 仅需万级别的数据量即可达到或超越 LANs 的准确性。此外，作为一种生成模型，MNLE 不仅能够估计似然值，还能够充当模拟器，可以迅速生成合成数据，从而避免了实际模拟运算的需求。这一特性在评估模型性能和计算模拟器内部信息(如参数的梯度信息)方面显示出潜在优势。更重要的是，MNLE 已被整合至 Python 开源工具包“sbi”中，这为在认知神经科学领域中实现无似然推断方法提供了基础(Tejero-Cantero et al., 2020)。

Bayesflow Radev 等人(2020)首次将神经后验估计方法应用于认知建模领域，提出了 BayesFlow 框架，该框架结合了标准化流和分摊策略，旨在通过一次性训练实现对认知模型参数的快速和高效推断(Radev et al., 2022; Radev, Voss, et al., 2020)。

与 MNLE 的不同之处在于，BayesFlow 引入了摘要统计网络(summary statistics network)，即$\mathcal{H}_\eta$，以解决基于深度学习的近似贝叶斯计算不允许试次数量变化的问题，该网络可以将任意长度的数据$D^{1:M}$(假设这些数据来自于独立同分布，并且具有置换不变性)映射为固定长度的嵌入向量$\mathcal{H}_\eta(D)^{1:K}$，其中$K$代表摘要统计维度的潜在维度。因此，参数后验的标准化流并不是直接基于数据进行变换，而是基于数据的摘要统计和参数先验进行可逆变换，即$z_D = f(\theta; \mathcal{H}_\eta(D))$和$\theta = f^{-1}(z_D; \mathcal{H}_\eta(D))$ (Radev et al., 2022)。

Bayesflow 执行无监督学习任务，所需的训练数据集形式为$\left\{\tilde{\theta}_i, \widetilde{D}_i^{1:M_i}\right\}_{i=1}^{N}$，其中$\widetilde{D}_i^{1:M_i}$用于生成摘要统计$\mathcal{H}_\eta(D)^{1:K}$。需要注意的是，每对数据的样本数量$M_i$存在差异，这使得训练后的模型能够适应各种真实样本数量。在训练期间，$\left\{\tilde{\theta}_i, \mathcal{H}_\eta(D)^{1:K}\right\}$共同输入条件可逆神经网络生成$z_D$，之后通过最小化 KL(Kullback-Leibler)散度，即$\underset{\phi,\eta}{min}\mathbb{KL}(p(\theta|D)||p_\phi(\theta|\mathcal{H}_\eta(D)))$，共同优化摘要统计和条件可逆神经网络两个网络，实现对于真实后验的近似(Radev et al., 2022)。

证据网络训练特点 Radev 等人(2021)首次在认知建模领域中使用神经模型比较方法，聚焦于解决贝叶斯模型比较(Bayesian model comparision，BMC)问题。他们提出证据网络(evidential network)，借鉴神经似然比估计的思想，可以直接计算近似贝叶斯因子(Bayes factor, BF)或模型后验概率(posterior model probabilities, PMP)，并可扩展至对于多个模型的比较。该方法完全基于模拟，具有分摊性质，并且无需为各模型训练独立神经网络。

证据网络训练的关键在于准备训练数据$\{\mathcal{M}_j, \widetilde{D}_i^{1:M_i}\}_{i=1}^{N}$：首先根据模型先验$p(\mathcal{M}_j)$采样得到模型$\mathcal{M}_j$；随后根据模型对应的参数先验分布$p(\theta_i|\mathcal{M}_j)$采样得到参数$\theta_i$；最后再根据参数生成数据$\widetilde{D}_i^{1:M_i}$，其中数据的样本量为$1{:}M$。证据网络执行监督学习分类任务，证据网络接受输入数据$\widetilde{D}_i^{1:M_i}$，并输出模型的后验概率$p(\mathcal{M}_j \mid D)$，可表示为$f_\phi: D \to p(\mathcal{M}_j \mid D)$，其中输出为$J$维向量，对应每个模型出现的概率。值得注意的是，为了量化神经网络估计的不确定性，证据网络不是直接输出，而是作为狄利克雷分布的浓度参数，结合模型先验来形成二阶概率分布。因此，证据网络的输出可被视为绝对证据，而狄利克雷分布的二阶概率分布可用于定量描述神经网络估计的不确定性(Radev et al., 2021)。

基于神经模拟推断的认知建模方法

Neural simulation-based inference: a neural network and simulation-based inference approach to cognitive modelling

Wanke Pan

Hu Chuan-Peng

(School of Psychology, Nanjing Normal University,

Jiangsu, Nanjing 210097, China)

Abstract

Cognitive modeling involves instantiating theoretical or model-based knowledge of cognitive processes into computational models and validating these theories by fitting behavioral and/or neuroimaging data. It enhances our understanding of human cognition through quantitative analysis and aids in the study of individual differences. Despite the ability of theory-driven computational models to generate simulated data, their complexity makes it difficult to determine the likelihood function, hindering the estimation of model parameters and comparisons between models based on observed data. This challenge is known as the inverse problem of generative modeling, and in response to the dilemma that the likelihood function is impossible or difficult to compute, which in turn gave rise to likelihood-free inference.

likelihood-free inference, a case of simulation-based inference, which uses simulated data to approximate or circumvent the calculation of the likelihood function. This method enables the fitting and evaluation of the model. Approximate Bayesian Computation (ABC) and Probability Density Approximation (PDA) are two common techniques used in likelihood-free inference. ABC compares simulated data to observed data, while PDA uses a kernel density estimator to approximate the likelihood function. Both methods are powerful tools for model selection and parameter estimation, particularly when the likelihood function is intractable or unknown. However, these methods may suffer from the curse of dimensionality, as the number of model parameters increases, the computational cost and complexity of the simulation-based approaches grow exponentially, which can make the methods impractical for high-dimensional data and models.

Advances in deep learning and neural network technologies have led to the emergence of a new neural simulation-based inference (NSBI). NSBI leverages the power of deep learning to

address the limitations of traditional simulation-based methods. NSBI possesses an amortization property, which allows the generation of a large amount of simulated data to be integrated into the training process. Consequently, once the neural network is trained, the inference step can be performed without the need for generating additional simulated data, significantly reducing computational overhead. Furthermore, with the advancement of neural conditional probability density estimation techniques, such as Normalizing Flows, NSBI can easily train models to estimate likelihood functions and posterior distribution functions. Neural posterior estimation methods, which offer a departure from conventional ABC techniques, directly sample from the posterior distribution using input data. Similarly, neural likelihood estimation methods, a substitute for PDA, compute likelihood values by leveraging both input data and prior parameters. These techniques enable the construction of flexible and complex cognitive models, making NSBI a promising approach for likelihood-free inference in high-dimensional settings.

NSBI has gained significant attention in the field of cognitive modeling and has been applied to various scenarios, including likelihood estimation, posterior estimation, and model comparison. For instance, neural likelihood networks such as LANs and MNLE are employed to estimate the likelihood function for intricate cognitive models, with MNLE being particularly adept at handling mixed data types with both continuous and discrete variables. In parameter inference for cognitive models, Bayesflow stands out as a neural posterior network capable of tackling a diverse array of complex models. For model comparison, Evidence Networks and Hierarchical Evidence Networks are utilized, with the Hierarchical variant being well-suited for nested data structures. Tools such as LANs, sbi, and Bayesflow are continuously optimizing the workflow of neural network-based simulation inference. These advancements have enabled NSBI to be applied to large-scale studies involving over a million data points, allowing for the construction of complex models that were previously difficult to build. This approach not only facilitates rapid validation of models and theories but also helps identify flaws and optimize performance, making NSBI a transformative technique with the potential to revolutionize the understanding and analysis of intricate systems.

While these new technologies show promise in cognitive modeling, this paper also discusses their limitations and offers constructive guidance for their use. The paper discusses the trade-offs of neural network training costs and model accuracy, the integration of NSBI methods for comprehensive analysis, and the importance of effective training for reliable models. It also emphasizes the potential for these technologies to enhance interdisciplinary collaboration, particularly between neuroscience, cognitive modeling, and other fields, to deepen our understanding of the mind and brain.

Keywords: Cognitive Modeling, Generative Models, Bayesian, Simulation-based Inference, Neural Network